

WHAT IS CLAIMED IS:

1 1. A method of predicting aggregate behavior of a population, the method
2 comprising:

3 providing a modeling system configured to model aggregate behavior of a
4 population as a function of aggregate on-line interest data, the on-line interest data based on
5 passive observation of on-line behavior of a subpopulation, wherein the on-line behavior is
6 related to, but different than, the behavior to be modeled, and wherein the subpopulation
7 comprises a subset of the population;

8 inputting to the modeling system on-line interest data related to a subject;
9 generating, with the modeling system, a prediction of aggregate behavior
10 related to the subject.

11 2. The method of claim 1 wherein the modeling system is further
12 configured to model aggregate behavior of the population as a function of characteristics of
13 the subject to which the aggregate behavior is related, the method further comprising
14 inputting to the modeling system data related to characteristics of the subject.

15 3. The method of claim 1 further comprising training the modeling
16 system with a learning data set, the learning data set including:

17 on-line interest data related to another subject, the another subject related to
18 the subject; and
19 actual aggregate behavior data relating to the another subject.

20 4. The method of claim 1 wherein the on-line interest data includes on-
21 line usage data.

22 5. The method of claim 1 wherein the aggregate behavior to be modeled
23 is aggregate economic activity.

24 6. The method of claim 5 wherein the aggregate economic activity to be
25 modeled is related to a product.

26 7. The method of claim 6 wherein the product is selected from the group
27 consisting of a movie, a video tape, a CD, a DVD, a model of automobile, a book, a toy, an
28 appliance, an electronic device, a pharmaceutical product, and a software product.

1 8. The method of claim 5 wherein the aggregate economic activity to be
2 modeled is related to a service.

1 9. The method of claim 5 wherein the aggregate economic activity to be
2 modeled is related to a financial security.

1 10. The method of claim 1 wherein the aggregate behavior to be modeled
2 is an extent of a disease.

1 11. A system for predicting aggregate behavior of a population, the system
2 comprising:

3 a modeling system configured to model aggregate behavior of a population as
4 a function of aggregate on-line interest data, the on-line interest data based on passive
5 observation of on-line behavior of a subpopulation, wherein the on-line behavior is related to,
6 but different than, the behavior to be modeled, and wherein the subpopulation comprises a
7 subset of the population; and

8 a module for receiving on-line interest data related to a subject and providing
9 the on-line interest data to the modeling system;

10 wherein the modeling system generates a prediction of aggregate behavior
11 related to the subject using the on-line interest data.

1 12. The system of claim 11 wherein the modeling system is further
2 configured to model aggregate behavior of a population as a function of characteristics of the
3 subject to which the aggregate behavior is related, the system further including a module for
4 receiving data related to characteristics of the subject and providing the data related to
5 characteristics of the subject to the modeling system.

1 13. The system of claim 11 further including a training module that trains
2 the modeling system with a learning data set, wherein the learning data set includes:

3 on-line interest data related to another subject, the another subject related to
4 the subject; and

5 actual aggregate behavior data relating to the another subject.

1 14. A method of training a modeling system to predict aggregate behavior
2 of a population, the method comprising:

3 providing a modeling system;
4 providing a learning data set including:
5 actual aggregate behavior data related to a first subject; and
6 aggregate on-line interest data related to the first subject, the on-line
7 interest data based on passive observation of on-line behavior
8 of a subpopulation, wherein the on-line behavior is related to,
9 but different than, the actual behavior, and wherein the
10 subpopulation comprises a subset of the population;
11 training the modeling system with the learning data set to minimize the error
12 between a predicted aggregate behavior related to the first subject generated by the modeling
13 system and the actual aggregate behavior related to the first subject.

14
15. The method of claim 14 wherein the learning data set further includes:
16 actual aggregate behavior data related to a second subject related to the first
17 subject; and

18 aggregate on-line interest data related to the second subject, the on-line
19 interest data related to the second subject based on passive observation of on-line behavior of
20 the subpopulation, wherein the on-line behavior is related to, but different than, the actual
21 behavior;

22 wherein training the modeling system with the learning data set includes
23 minimizing the mean-square error between the predicted aggregate behavior related to the
24 first subject generated by the modeling system and the actual aggregate behavior related to
25 the first subject and between a predicted aggregate behavior related to the second subject
26 generated by the modeling system and the actual aggregate behavior related to the second
27 subject.

28
1 16. A method of predicting a measure of aggregate economic activity
2 related to a product, the method comprising:
3 providing a modeling system configured to model aggregate economic activity
4 of a type of product as a function of aggregate on-line interest data related to products
5 comprising the type, wherein the on-line interest data is based on passive observation of on-
6 line behavior of a subpopulation, wherein the on-line behavior is related to, but different than,
7 the economic activity to be modeled, and wherein the subpopulation comprises a subset of a
8 population that engages in the economic activity to be modeled;

generating a prediction of the measure of aggregate economic activity related to the first product with the modeling system.

1 17. The method of claim 16 wherein the modeling system is further
2 configured to model aggregate economic activity of the type of product as a function of
3 characteristics of products comprising the type, the method further comprising inputting to
4 the modeling system data related to characteristics of the first product.

on-line interest data related to a second product comprising the type; data related to characteristics of the second product; and aggregate economic activity data relating to the second product.

1 19. The method of claim 18 wherein training the model includes:
2 adding to the learning data set additional data related to characteristics of the
3 second product; and
4 retraining the modeling system with the learning data set.

1 21. The method of claim 20 wherein training the model includes:
2 adding to the learning data set additional on-line interest data related to the
3 second product; and
4 retraining the modeling system with the learning data set.

1 23. The method of claim 16 wherein the on-line interest data related to the
2 first product includes counts of search queries at a web site that include a phrase related to the
3 first product.

1 24. The method of claim 16 wherein the on-line interest data related to the
2 first product includes an on-line interest measurement provided by a web site.

1 25. The method of claim 24 wherein the on-line interest measurement
2 provided by a web site is a fictional stock price of the first product.

1 26. The method of claim 24 wherein the on-line interest measurement
2 provided by a web site is a percentage of users of the web site initiating searches related to
the first product.

1 27. The method of claim 16 wherein the on-line interest data related to the
first product includes aggregate Internet usage data related to the first product.

1 28. The method of claim 27 wherein the aggregate Internet usage data
related to the first product includes statistics based on analyses of online events related to the
first product.

1 29. The method of claim 28 wherein online events include a result of a
2 client making a request of a server and the server providing a response to the client.

1 30. The method of claim 28 wherein the analyses of online events
2 includes:
3 automatically associating each online event with one or more subjects;
4 accumulating counts for events by subject; and
5 outputting the accumulated counts for each subject.

1 31. The method of claim 30 wherein the analyses of online events further
2 includes:
3 identifying one or more categories relevant to each subject;
4 accumulating counts for events by category; and
5 outputting the accumulated counts for each category.

1 32. The method of claim 30 wherein the analyses of online events further
2 includes determining if a subject for an event is a canonical equivalent of another subject; and
3 wherein counts for canonical equivalents are accumulated together.

1 33. The method of claim 30 wherein the analyses of online events further
2 includes normalizing counts for events over a field of events, and wherein outputting the
3 accumulated counts includes outputting the normalized counts.

1 34. The method of claim 30 wherein the analyses of online events further
2 includes:

3 determining a set of one or more demographic parameters relating to users that
4 prompt the events; and

5 using the set of one or more demographic parameters to partition the counts by
6 demographic divisions.

1 35. The method of claim 16 wherein the first product is selected from the
2 group consisting of a movie, a video tape, a CD, a DVD, a model of automobile, a book, a
3 toy, an appliance, an electronic device, a pharmaceutical product, and a software product.

1 36. The method of claim 16 wherein the predicted measure of aggregate
2 economic activity is a predicted number of sales during a period of time.

1 37. The method of claim 16 wherein the predicted measure of aggregate
2 economic activity is a predicted monetary value of sales during a period of time.

1 38. A system for predicting a measure of aggregate economic activity
2 related to a product, the system comprising:

3 a modeling system configured to model aggregate economic activity of a type
4 of product as a function of aggregate on-line interest data related to products comprising the
5 type, wherein the on-line interest data is based on passive observation of on-line behavior of a
6 subpopulation, wherein the on-line behavior is related to, but different than, the economic
7 activity to be modeled, and wherein the subpopulation comprises a subset of a population that
8 engages in the economic activity to be modeled; and

9 a module for receiving on-line interest data related to a first product
10 comprising the type and providing the on-line interest data to the modeling system;

11 wherein the modeling system generates a predicted measure of economic
12 activity related to the first product using the on-line interest data.

1 39. The system of claim 38 wherein the modeling system is further
2 configured to model aggregate economic activity of the type of product as a function of
3 characteristics of products comprising the type, the system further including a module for
4 receiving data related to characteristics of the first product and providing the data related to
5 characteristics of the first product to the modeling system.

1 40. The system of claim 39 further including a training module that trains
2 the modeling system with a learning data set, wherein the learning data set includes:

3 on-line interest data related to a second product comprising the type;
4 data related to characteristics of the second product; and
5 aggregate economic activity data related to the second product.

6 41. The system of claim 38 further including a training module that trains
7 the modeling system with a learning data set, wherein the learning data set includes:

8 on-line interest data related to a second product comprising the type; and
9 aggregate economic activity data related to the second product.

10 42. The system of claim 38 further comprising an aggregate Internet usage
11 statistics generator that provides aggregate Internet usage statistics related to the first product
12 to the module for receiving on-line interest data.

13 43. The system of claim 42 wherein the aggregate Internet usage statistics
14 generator includes:

15 an activity input for receiving data related to events on a set of servers;
16 means for categorizing events into categories;
17 means for associating events with subjects, wherein counts are maintained for
18 each subject and wherein subjects are associated with categories;
19 a normalizer for normalizing counts for events over a field of events; and
20 a result output for outputting results of the normalizer as the online usage
21 statistics.

22 44. A method of training a modeling system to predict aggregate economic
23 activity related to a product comprising a type of products, the method comprising:

3 providing a modeling system;
4 providing a learning data set including:
5 an actual measure of aggregate economic activity related to a first
6 product comprising the type; and
7 aggregate on-line interest data related to the first product, the on-line
8 interest data based on passive observation of on-line behavior
9 of a subpopulation, wherein the on-line behavior is related to,
10 but different than, the actual economic activity, and wherein the
11 subpopulation comprises a subset of a population that engages
12 in the economic activity;
13 training the modeling system with the learning data set to minimize the error
14 between a predicted measure of aggregate economic activity related to the first product
15 generated by the modeling system and the actual measure of aggregate economic activity
16 related to the first product.

1 45. The method of claim 44 wherein the learning data set further includes:
2 an actual measure of aggregate economic activity related to a second product
3 comprising the type;

4 aggregate on-line interest data related to the second product, the on-line
5 interest data based on passive observation of on-line behavior of the subpopulation, wherein
6 the on-line behavior is related to, but different than, the actual economic activity;
7 wherein training the modeling system with the learning data set includes
8 minimizing the mean-square error between the predicted measure of aggregate economic
9 activity related to the first product generated by the modeling system and the actual measure
10 of aggregate economic activity related to the first product and between the predicted measure
11 of aggregate economic activity related to the second product generated by the modeling
12 system and the actual measure of aggregate economic activity related to the second product.